Workgroup 5: PT Provider Assessment of Laboratory Performance

Mr. Daniel Tholen

WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF PT/EQA PROGRAMS THAT ARE MANAGED FOR EDUCATIONAL AND/OR REGULATORY PURPOSES?

- Regulatory PT assures total participation
- Voluntary PT provides incentive of peer recognition, but does not assure total participation
- Regulatory PT provides more comprehensive oversight
- Educational component needed in both schemes; education should be the primary focus of PT
- Type of challenges offered for PT/EQA often driven by educational vs regulatory schemes.
- Punitive aspect of regulatory PT discourages difficult PT challenges

ARE THERE SOME WAYS TO SATISFY BOTH SETS OF NEEDS?

- Have separate regulatory and educational PT challenges
- If laboratory given opportunity to take corrective action before sanctions applied, regulatory and educational PT could be the same

WHAT ARE THE CONSIDERATIONS FOR DETERMINING WHEN TEST PERFORMANCE CAN BE GRADED?

- When to grade is dependent on the method for determining the target value and range
- If the program is purely educational, performance may not need to be graded

SHOULD EVERY TEST BE GRADED, OR ARE THERE SOME TESTS FOR WHICH PT/EQA IS PREMATURE, REDUNDANT, OR UNNECESSARY?

- Some PT challenges should be educational – may be evaluated for educational needs but not graded for actual performance
- Do not grade when
  - there is a lack of consensus
  - results cannot be grouped into all method mean group and peer groups are too small
  - there is a known method bias
  - starting a new PT program for new method or analytes

ARE THERE OBJECTIVE CRITERIA FOR MAKING THIS DECISION?

- Difficult to determine objectively – assessment tools are often subjective
- Objective criteria could be based on
  - the variability of the target
  - extent of experience with a new PT analyte, method, or program

WHAT PERFORMANCE MEASURES ARE APPROPRIATE FOR QUANTITATIVE AND QUALITATIVE TESTS?

- Accuracy - when results are traceable to reference materials/reference methods
- Interlaboratory consensus/variation and reference laboratory results.

Comment: Approaches can be used for both qualitative and quantitative tests, but no group agreement on number of reference laboratories which must agree, or the degree of consensus required, for an answer to be considered “true.”

WHAT STATISTICAL TOOLS CAN BE USED TO MEASURE THESE CHARACTERISTICS?

- The group did not possess strong expertise in this area, and preferred not to contribute disinformation to the conference

WHAT CHARACTERISTICS OF PT/EQA PERFORMANCE CAN BE EVALUATED?

- Test-inherent problems, such as information about outdated methods, varying performance characteristics of different methods and instruments, etc.
- Lab-inherent problems, such as the effects of different levels of training, technologist competence, problems with quality laboratory management, etc.

IS IT POSSIBLE TO EVALUATE A LABORATORY'S INTERPRETATION OF TEST RESULTS? – YES!

- Evaluation can be oriented towards either the analytical process, clinical interpretation of results, or both.
- Evaluation of interpretation skills is best considered educational, not regulatory. Can be assessed using questions on clinical practice, simulated patients, etc. Reference answers are likely to be a consensus of experts, rather than provable outcomes.

Additional Comments

1. It is difficult to do a really “blind” PT challenge.
2. We should allow a lab to flag test results it does not trust:
   a) Could be a lab problem, such as a new test that the lab is not yet offering to patients.
   b) Could be a problem that the lab believes is present in the PT material, such as degradation during shipping, etc.
3. Manufacturers of diagnostic products may find it useful to participate in PT/EQA challenges.

HOW SHOULD PERFORMANCE GOALS FOR LABORATORIES BE DETERMINED?

- Establish target values and limits by:
  - Reference methods
  - Weighed-in values
  - Reference labs
  - Peer groups (last resort)
- Can use a combination of the above
- Goals differ in different situations, and may be dependent on the difference between PT specimens and patient samples (commutability)

SHOULD PERFORMANCE BE MEASURED RELATIVE TO OTHER LABORATORIES, OR WITH OBJECTIVE GOALS?

- Performance goals should always be based on clinical relevance
- PT samples should be designed with values around critical decision points

WHAT ARE THE CONSIDERATIONS FOR DETERMINING THAT A LABORATORY'S PERFORMANCE IS *ACCEPTABLE*?

- Acceptable performance is determined by:
  - Accrediting bodies
  - Local regulators
  - Professional societies
  - Expert opinion
- Factors to consider:
  - QC programs
  - Accuracy / Reproducibility / Precision
  - On time results
  - Clerical errors

WHAT ARE THE CONSIDERATIONS FOR DETERMINING THAT A LABORATORY'S PERFORMANCE IS *UNACCEPTABLE*?

- Unacceptable performance is influenced by:
  - Level of available technology
  - Lab infrastructure
  - Level of personnel training

SHOULD IT BE BASED ON A SINGLE TEST RESULT, A SET OF RESULTS IN A SINGLE TEST EVENT, OR RESULTS ACROSS SEVERAL TEST EVENTS?

- Measure unacceptable performance by trending:
  - Across samples
  - Across PT events

SHOULD PERFORMANCE GOALS BE THE SAME FOR ALL TYPES OF LABORATORIES?

- Ideally, performance goals should be the same for all types of laboratories, because patients deserve reliable results regardless of the setting; but test complexity should be considered
- Screening vs. confirmatory results
- Qualitative vs. quantitative (unresolved issue)

WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF "BLIND" PT/EQA, WHERE SAMPLES ARE INTRODUCED INTO THE LABORATORY AS PART OF THE NORMAL WORKLOAD INSTEAD OF BEING IDENTIFIED AS PT/EQA SAMPLES?

- Advantages
  - Would test pre- and post-analytical, as well as analytical, processes
  - More feasible to accomplish internal PT/QA
  - Potential special role in cases of alleged poor quality or falsification
  - More plausible for specific application (eg. newborn screening)

WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF "BLIND" PT/EQA, WHERE SAMPLES ARE INTRODUCED INTO THE LABORATORY AS PART OF THE NORMAL WORKLOAD INSTEAD OF BEING IDENTIFIED AS PT/EQA SAMPLES?

- Disadvantages
  - High overhead (cost, labor, administration)
  - Legal issues (false pt records, anonymization)
  - For resource challenged countries, primary focus should be on traditional PT

WHAT ARE THE CRITERIA THAT SHOULD BE USED TO DETERMINE THE FREQUENCY OF PT/EQA?

- Stability of process
  - Analytical stability, staffing
- Cost
- Purpose
  - Regulatory vs QA
- Turnaround-time of EQA provider feedback
- Transportation/distribution
- Workload planning (more frequent, fewer samples are preferred)

WHAT FACTORS SHOULD BE CONSIDERED IN DEFINING PEER GROUPS IF SUCH GROUPS ARE USED IN DETERMINING ACCEPTABLE LABORATORY PERFORMANCE?

A. Sample behavior: differs for different analyte / sample types

B. Specialties: do not peer group all specialties the same way, e.g., Chemistry, Hematology, Coag.

C. Application: grouping needs can be affected by the use of test (direct to patients or reference laboratory, developed country)

D. Size of peer group.

WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF LABORATORIES RECEIVING IDENTICAL CHALLENGES IN EVERY TEST EVENT?

Three ways to view question:

1. Two of the Same sample within shipment.
   - Can check on repeatability within lab.
2. Same sample in different shipments.
   - Allow checking target values.
   - Can track effectiveness of training.
   - Allow seeing change in performance
3. **Same sample in a shipment to all laboratories.**

WHAT ARE THE ADVANTAGES AND DISADVANTAGES OF LABORATORIES RECEIVING IDENTICAL CHALLENGES IN EVERY TEST EVENT?

Advantages:

- Inhibits cheating
- Maximum peer group size
- Easier logistically
- Fewer samples needed

Disadvantages:

- Participants can recognize patterns
- Can limit variety of sample types and range tested
- Handling of rare disease could cause outbreak.

Summary

- All laboratories should be required to participate in PT/EQA, ***but…***
- The primary focus should be education
- Grades should be based on clinically relevant criteria
- Performance measures can include accuracy, precision, and interpretation
- Reference values are preferred to consensus

Summary

- Laboratory performance cannot be judged solely on PT/EQA results
- PT performance should be based on trends over different samples and events, not single tests
- No clear advantages of blind PT
- Peer grouping criteria can vary by situation
- Because evaluations should be based on clinical relevance, there could be different grading criteria for different laboratory settings

SUBGROUP LEADERS AND RECORDERS

- Dev Howerton
- Tim O'Leary
- Leigh Dini
- Elizabeth Melnyk
- Daniel Edson
- Adam Manesterski
- Mary Kimberly
- Darshan Singh
- Rex Astles
- John Hancock